Neural Networks for Diagnosing Skin Cancer and their Resulting Inequities

Shane Nelson

One in five Americans will develop skin cancer by the age of 70 and more people are diagnosed with skin cancer than all other cancers combined.[1] Increases in melanoma incidence, the most deadly skin cancer, are predicted, compounding these statistics. Luckily, melanoma and other skin cancers are highly treatable if diagnosed early. The five-year survival rate for melanoma increases from 14% if caught in the late stages to over 99% if the cancer is caught early.[2] These facts have motivated extensive research into machine learning classification algorithms to quickly and easily identify dangerous lesions. If scalable, algorithms can be deployed remotely on smartphone applications, increasing access to potentially lifesaving diagnoses. Promisingly, these technologies could provide universal access to patients who may otherwise avoid healthcare for reasons such as lack of insurance or healthcare related anxiety. Computer-aided diagnostic tools are not new but have seen improvements in recent years.

Prior to advances in black-box computer vision algorithms such as convolutional neural networks, computer aided diagnostic tools relied on classification algorithms such as logistic regression, k nearest neighbors, and tree based methods. Algorithms such as these took precleaned and filtered dermoscopic images, as well as features such as descriptions of the borders of lesions (a classical diagnostic tool to recognize malignant lesions) and outputted probability functions of group membership, often malignant or benign. These methods typically performed at the level of trained physicians[3] and their decisions were traceable due to their non-black-box nature. As could be guessed by the development of newer technologies, these methods had their flaws, namely, they required extensive pretreatment to images and complicated feature selection algorithms. Additionally, even after these steps, they underperformed at discerning rarer malignancies, such as melanomas, due to their relative infrequency in training datasets.

[1] "Skin Cancer Facts & Statistics." *The Skin Cancer Foundation*, 6 Feb. 2024
[2] Esteva, Andre, et al. "Dermatologist-level classification of skin cancer with deep neural networks." *Nature*, vol. 542, no. 7639, 2017, pp. 115–118,
[3] Masood, Ammara, and Adel Ali Al-Jumaily. "Computer Aided Diagnostic Support System for Skin Cancer: A Review of Techniques and Algorithms." *Int J Biomed Imaging*, vol. 2013, 2013, pp. 323268–22

Recently, a new state-of-the-art procedure has been developed, a computer vision algorithm called the convolutional neural network (CNN). The CNNs used as diagnostic tools are specialized deep learning algorithms that take image pixel data as their input and output probability functions. The output functions give probabilities of group membership over a set of possible groups. Commonly these group sets begin as benign and malignant and are further divided into subsets describing precise types of skin lesions.[2] Convolutional neural networks work well for this application because they are good at classifying images and do not require pretreatment of the input images. This fact allows CNNs to be trained on the large, untreated, labeled datasets which have recently become available. While these models allow for the possibility of increased access to care and early detection rates, they are not without ethical concerns.

Virtually all computer aided diagnostic tools are reflections of the datasets they are trained on. Because of this, any biases against different groups or disproportionate representation of different groups present in training datasets will be reflected in differing accuracy of predictions made using the datasets. This problem is rampant in computer vision algorithms created to diagnose skin cancer.[4] This is largely due to the vast underrepresentation of images of lesions on Black patients in the presently available datasets. Because these models are not trained on images of Black patients they unsurprisingly perform worse on Black patients.[5]

Training set underrepresentation has many causes. Some of the studied causes are worse access to healthcare, lower rates of skin cancers in patients of color, and the fact that practices with less resources are less likely to contribute to medical datasets.[6] Underrepresentation is especially concerning considering the racial inequities associated with skin cancer diagnosis and prognosis.

Black patients have only a 70% five-year melanoma survival rate compared to 94% for white patients.[7] This is due to a multitude of factors, briefly: increased proportion of late-stage diagnoses, worse access to healthcare, poorer medical training for recognizing skin conditions in Black patients versus white patients[8], and systemic racism. There are also critical differences in

---

[4] Seyyed-Kalantari, Laleh, et al. "CheXclusion: Fairness gaps in deep chest X-ray classifiers." *Pac Symp Biocomput*, vol. 26, 2021, p. 232.
[5] Dildar, Mehwish, et al. "Skin Cancer Detection: A Review Using Deep Learning Techniques." *Int J Environ Res Public Health*, vol. 18, no. 10, 2021, p. 5479.
[6] Hing E, Burt CW. Are there patient disparities when electronic health records are adopted? J Health Care Poor Underserved. 2009 May;20(2):473-88. doi: 10.1353/hpu.0.0143. PMID: 19395843.
[7] "Skin Cancer in People of Color." *Office of Student Life*, 20 Oct. 2023
[8] Frazier, Winfred Taylor, et al. "Common Dermatologic Conditions in Skin of Color." *Am Fam Physician*, vol. 107, no. 1, 2023, pp. 26–34.

how skin cancer presents in Black patients. Typically, cancers are found in harder to notice areas such as the bottom of the foot or under the fingernails.[7] These differences, along with discrepancies in medical training to recognize them, lead to Black patients being over three times as likely to be diagnosed in the later stages when compared to white patients.[7] While skin cancer incidence is higher for white Americans, these factors combine to make the prognosis worse for Black patients on average. A tool which can increase the probability of early diagnosis for patients without insurance or access to high quality healthcare, both conditions which disproportionately affect non-white Americans, would go an extremely long way in correcting the racial inequities in prognosis for skin cancers. However, this technology is currently inaccessible to patients of color.

The underrepresentation of Black patient images in training datasets only serves to widen racial gaps in healthcare if computer vision becomes the gold standard for early detection of skin malignancies. While not a perfect proxy to skin lesion classification algorithms, there has been extensive research into mitigating accuracy biases in CNNs used to diagnose conditions using X-rays. These algorithms suffer from significant discrepancies in accuracy based on the sex of the patient.

Possibly the most obvious solution to the accuracy discrepancy is to increase the amount of images of Black patients in the training image sets. While this could go a long way, there is reason to believe that increasing representation alone may not be enough to solve the problem. One study which examined the relative accuracy of diagnostic X-ray CNNs found that the disparity between male sexed and female sexed patients was not correlated with the subgroup's proportional disease prevalence. This finding suggests that "[increased] dataset membership cannot always ameliorate biases."[4] Once again this is not a perfect proxy, but it does advise that simply increasing the proportion of images of Black patients alone may not be enough to counteract the biases reflected in diagnostic CNNs.

Another consideration which has been proposed and researched is embedding a respect for fairness into the algorithms themselves, much like the examples of college admissions and loan decision making which were discussed in *The Ethical Algorithm*. This type of correction works by requiring that the algorithm balances its success or error rates, ensuring that each protected group sees similar benefits. For a skin cancer classification algorithm this could mean that the proportion of true positives (algorithm classifies skin cancer, doctors confirm skin cancer) to positives (algorithm classifies skin cancer) is balanced for every racial group.[9] However, this is likely not possible in any relevant capacity currently due to the lack of representation of patients of color in training datasets. That is, any algorithm which is optimized for any significant measure of fairness would be likely to suffer large inaccuracies. This reason and the findings in X-ray CNNs suggest that both increased dataset representation and algorithm design with respect to fairness are necessary in order for skin cancer diagnostic algorithms to ameliorate racial healthcare disparities, or at the very least not exacerbate them.

Technologies such as machine-learning-based diagnostic tools present an exciting opportunity for cheap early detection, with the potential to save thousands of lives yearly if embedded into accessible technologies such as smartphones. Algorithmic diagnostic aids are especially salient as the incidence rates of melanomas and other skin cancers are only expected to rise in the future. This is not without its challenges however. Algorithms such as these are trained on extensively majority-white datasets which prevents them from learning to diagnose skin cancers in patients of color. If these technologies were to become the most prevalent method for early detection without a consideration to fairness, the discrepancies they promote would only serve to widen pre existing healthcare inequities. This should be combated by increasing the representation of patients of color in training datasets and by embedding fairness into the algorithms themselves, similarly to how newer CNN diagnostic tools using X-rays have been developed.

[9] Chen, Irene Y, et al. "Ethical Machine Learning in Healthcare." *Annu Rev Biomed Data Sci*, vol. 4, no. 1, 2021, pp. 123–144, https://doi.org/10.1146/annurev-biodatasci-092820-114757.